International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:7, No:7, 2013

On the Learning of Causal Relationships between Banks in Saudi Equities Market Using Ensemble Feature Selection Methods

Adel Aloraini

Abstract—Financial forecasting using machine learning techniques has received great efforts in the last decide . In this ongoing work, we show how machine learning of graphical models will be able to infer a visualized causal interactions between different banks in the Saudi equities market. One important discovery from such learned causal graphs is how companies influence each other and to what extend. In this work, a set of graphical models named Gaussian graphical models with developed ensemble penalized feature selection methods that combine ; filtering method, wrapper method and a regularizer will be shown. A comparison between these different developed ensemble combinations will also be shown. The best ensemble method will be used to infer the causal relationships between banks in Saudi equities market.

Keywords—Causal interactions , banks, feature selection, regularizere.

I. Introduction

Stock market prediction applications have been widely investigated using machine learning and data mining techniques. Artificial Neural networks(ANNs) considered to be one of the successful predictive techniques used in stock market. ANNs have shown a great implication on prediction from historical time series stock market data such as for modeling and forecasting [15]. Hybrid techniques for stock market predictions have shown another aspect of good future predictions using nonlinear modeling techniques[7]. To improve the performance of predictive ANNs models , some studies incorporated prior knowledge to improve the prediction accuracy and found it much better than standard ANNs[10]. Evolutionary algorithms are also used in stock market predictions. Genetics algorithms have been used for feature discretization that is later fed to ANNs for stock market predictions [8]. A study by[9] has shown that using SVM outperforms the prediction accuracy of ANNs and case-based reasoning (CBR). However, the study has also reported that the sensitivity of the upper bound-C and the kernel parameter δ^2 in SVM play a central role in SVM prediction accuracy.

In this work we are concerned about similar applications to stock market but from another aspect. We are interested in unsupervised learning paradigm that is later used for inference. The work in this paper is centralized on how to learn the causal-effect relationships between different companies in Saudi equities market using machine learning of graphical models to revealing the hidden causality between banks based on historical data from Saudi stock market. The set of graphical models are penalized Gaussian graphical models in which to map the relationship between features , we asses how informative/sensitive are the predictors to a particular feature using developed *multi-layer feature selection methods*. The multi-layer feature selection methods consist of a filter method , a wrapper method , and a regularizer using L1-regression[6]. The next sections will be organized as follows; related work to the proposed methods for the graphical models learning, detailed sections about the methods developed , results and discussion, and the conclusion with the future work.

A.Aloraini is with the Department of Computer Science, Science and Arts college-Alrass, Qassim University, Saudi Arabia,e-mail: (see http://www.coc.qu.edu.sa/en/dep/docsc/a.oraini/Pages/default.aspx).

II. Related Work

Machine learning of graphical models are widely used in different applications to infer the hidden relationship between number of variables/features(p) across different size of samples(n). Machine learning of graphical models are used to infer the gene-regulatory networks from gene expression datasets, and [11] gives an excellent review on different set of graphical models learning techniques to infer cellular networks and gene-regulatory networks. Machine learning of graphical models are also proposed to be important techniques in stock market such as using ANNs which are considered to be a graphical model representation in a compact way. Learning graphical models with embedded feature selection techniques also appear to be important due to the noise that is usually associated with stock market data collection. Feature selection techniques differ from each other depends on the way they incorporate in the model selection. They can be organized into three categories : filter methods, wrapper methods, and embedded methods. Filter methods assess the relevant features independently from the machine learning algorithm which is known to be a drawback of such feature selection methods[13]. Wrapper methods in the other hand, incorporate the feature selection tasks from within the machine learning algorithm and hence such learning considered mostly as a subset model selection since the feature selection task is wrapped within the machine learning task. The third feature selection category is embedded techniques. L1-regression, and LARS methods, are well known form of embedded techniques from which the most optimal parameters and features are learned simultaneously [6], [14]. L1-regression and its variants such as *Lasso* estimate are widely used in learning graphical models using regularisation, variable selection or covariance selection for high-dimensional datasets, where the number of predictor

variables is much larger than the number of samples ($p \gg n$) [2], [12].

In this work, a more constrained regularisation is achieved to determine the best features. For this purpose, we are going to develop ensemble feature selection methods that is consist of a filter method,a wrapper method, and lasso estimate, in order to learn the best features(banks) that have causal relationships on a particular bank in the context of penalized Gaussian graphical models.

In the next section , we will detail the different combinations of these ensemble feature selection methods and proceed to evaluate each combined ensemble method to choose the one that gives best prediction accuracy which in turn is used to learn the causal relationships between banks in Saudi equities market.

III. Methods

In this section, we give a broaden discussion about the developed methods used in this work.

A. Gaussian graphical models and linear regression

In a previous work [1], we showed that the distribution of some experimental datasets can be approximated by a multivariate Gaussian distribution. This distribution is known to be decomposable [3] into a product of conditional distributions:

$$P(X_1, X_2, X_3, ...) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2), ...$$

Where the relationship between predictors and a target variable fits a linear regression model. Based on this, for each target variable in the dataset, we sought a subset of good predictors assuming a linear model.

$$\mu_i = \beta_{i0} + \sum_{j=1}^{p} \beta_{ij} x_{ij}$$

However, It is noted that the normal Gaussian graphical models focuses on variable selection and not parameter estimation βs, such that the subset of predictors is chosen and then the parameters for the learnt subset of predictors are determined by maximum likelihood (least squares). For simultaneous subset selection and parameter estimation, Lasso estimate (L1-regression) is proposed recently [6], [14] which is able to find the subset of predictors and estimate βs in a more continuous way.

B. Lasso estimate (L1-regression)

Lasso is a shrinkage method in which many βs are 'shrunk' to zero [14]. This is because the penalty for large βs in lasso is very severe, being the sum of the absolute values of the regression coefficients(βs) $\sum_{j=1}^{p} |\beta_j|$ (in contrast to ridge regression $\hat{\beta}^{ridge}$ where the penalty is less strict $\sum_{j=1}^{p} \beta_j^2$). The lasso estimate $\hat{\beta}^{lasso}$ for the regression coefficients for a particular complexity parameter λ is:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{i=1}^{p} |\beta_j| \right\}$$

or equivalently (where λ is determined by s):

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2, \text{ subject to } \sum_{j=1}^{p} |\beta_j| \leq s$$

No penalty is applied to the intercept (β_0) , so $\beta_0 = \overline{y} = \sum_1^N y_i/n$, and the x_{ij} are centred.

Therefore , Lasso estimate is indeed an embedded feature selection method in which penalization is applied to choose the best subset of predictors by shrinking unimportant parameters(βs)=0.0. Hence , when it is applied to infer Gaussian graphical models, these graphical models are named *penalized Gaussian Graphical models*.

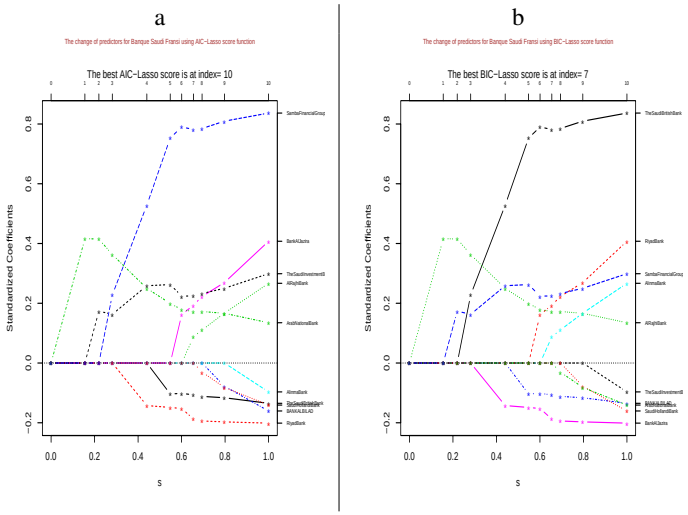

Fig. 1. (a) AIC scores all the resultant subset of predictors for Banque Saudi Fransi.(b) BIC scores all the resultant subset of predictors from Lasso for Banque Saudi Fransi bank.

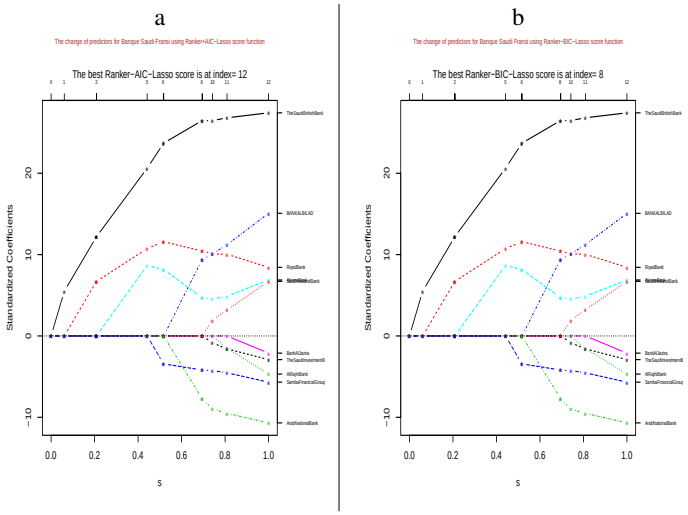

Fig. 2. (a) In this result, all predictors are ranked for Banque Saudi Fransi and then s is found using AIC-Lasso in(a) and s is found using BIC-Lasso in(b).

IV. A combination between a filter ,wrapper and lasso estimate feature selection methods to infer penalized Gaussian Graphical models

One way to select subset of predictors is to use wrapper methods that incorporate the feature selection tasks from

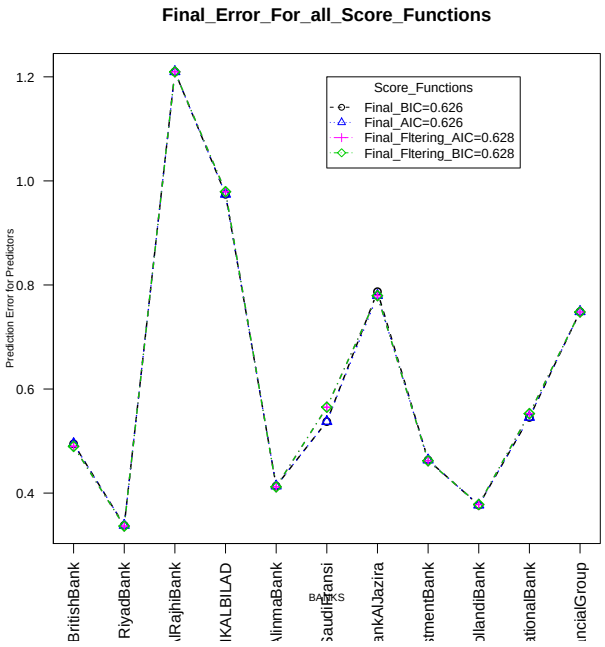

Fig. 3. This result shows the behaviour of prediction accuracy for each best subset of predictors for each bank from each ensemble feature selection method.

within the machine learning algorithm. For the work represented here , we will show how a search-score approach, as a wrapper method, can be joint within the lasso feature selection method to determine the best subset of predictors. We will use AIC and BIC score functions to justify the goodness of fit for the chosen penalized predictors and the proposed ensemble methods will be named AIC-Lasso, and BIC-Lasso ensemble feature selection methods. AIC and BIC score functions include a complexity penalty term that increases with the number of predictors(p):

$$\mathrm{AIC} = n\log(\mathrm{RSS}/n) + 2p$$

$$\mathrm{BIC} = n\log(\mathrm{RSS}/n) + p\log(n)$$

Where RSS is the residual sum of squares. Moreover, within the AIC-Lasso and BIC-Lasso we will inject more detailed steps to learn the *most* important predictors. Therefore, we will add an extra layer of feature selection called filter method [4] ,which ranks all the possible subset of predictors for each bank according to their importance using correlation coefficients(1) incrementally, from the highest to the lowest before using AIC/BIC-Lasso ensemble method for feature selection.

$$\mathrm{r} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (1)$$

The new ensemble method is named Ranker-AIC/BIC-Lasso ensemble method. For the sake of comparison between AIC-Lasso, BIC-Lasso, Ranker-AIC-Lasso, and Ranker-BIC-Lasso, we will evaluate them in terms of their prediction accuracy for the chosen subset of predictors. As a result, the best ensemble feature selection method will be used to learn a penalized graphical model to show how the causal-effect relationship between banks happens in Saudi equities market.

In the following section, we will show more details about the constructive proposed ensemble feature selection methods.

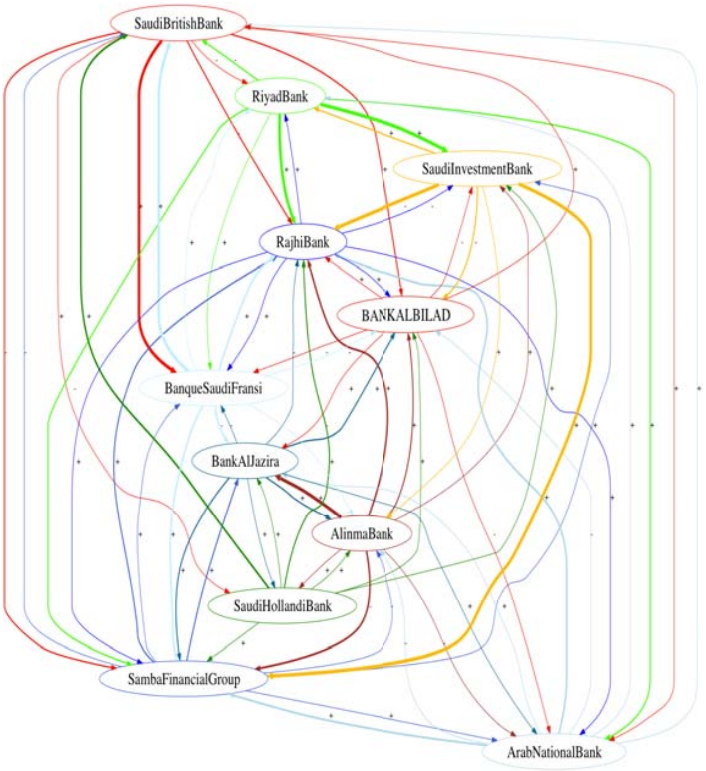

Fig. 4. The resultant causal gaphical model for banks in Saudi equities market from BIC-lasso ensemble method.

A. *Finding an optimal s using AIC/BIC-Lasso ensemble feature selection method.*

In Lasso, choosing the best subset of predictors is subject to choosing the best value for the tuning parameter s. For that purpose, we will use AIC/BIC score function to determine the optimal value of s. In learning the model, the best subset of predictors will be learnt and evaluated along with estimating βs.

The best subset of predictors will be chosen according to the most appropriate value of s that is chosen by AIC/BIC.

This is done by scoring all the models that are returned by different values of s. The model with the smallest AIC/BIC will be chosen as the best model. Therefore, the ensemble feature selection methods here can be seen as a combination between a wrapper feature selection method(search-score(AIC/BIC)) and an embedded feature selection(Lasso); hence named AIC/BIC-Lasso ensemble feature selection method. Fig.1(a),

and Fig.1(b) show how AIC-Lasso and BIC-Lasso are used to choose the best value of s, respectively for Banque Saudi Fransi.

B. Ranker-AIC-Lasso ensemble feature selection method

We also proposed a more detailed feature selection ensemble method embedded within the AIC-lasso/BIC-lasso ensemble methods. We injected more detailed steps to learn the *most* important predictors for each bank. All possible subset of predictors for each bank will be ranked according to the correlation coefficients incrementally, from the highest to the lowest, called feature ranking[4]. Following this, leave-one-out cross validation (LOOCV) is used to test the prediction error each time we remove a predictor from the set of predictors. The advantage of using the feature ranking method here within AIC-lasso/BIC-lasso score functions, is that when a subset of predictors is examined using only the AIC-lasso/BIC-lasso estimate, the optimal value of s is used to choose the best subset of predictors from all possible predictors *one time*. However, when feature ranking method is used we make several choices based on the ranked possible predictors. Each time we remove a predictor, we test how good are the remaining subset of predictors using AIC-lasso/BIC-lasso, and find the best s for this subset of predictors. We repeat the process until we test only the best predictor alone in the model. Intuitively, there is a clear advantage for adding feature ranking to the AIC/BIC-lasso estimate function that undertakes this process only one time. However, due to the greediness of ranking the predictors incrementally, it might not reach the optimal prediction accuracy comparing to AIC/BIC-Lasso feature selection methods which are considered as a less greedy search approach [6]. Fig.2(a), and Fig.2(b) shows how the best subset of predictors for Banque Saudi Fransi is chosen using Ranker-AIC/BIC-Lasso ensemble feature selection methods, and how the chosen predictors differ from those chosen by AIC-Lasso and BIC-Lasso in Fig.1(a), and Fig.1(b) for Banque Saudi Fransi.

V. Results and Discussion

In this section, an analytical comparison between AIC-lasso, BIC-lasso, Ranker-AIC-lasso, and Ranker-BIC-lasso ensemble feature selection methods, will be given. For evaluation, leave-one-out cross-validation is used (LOOCV), where the data is iteratively split to train and test (K=n). Finally, the average of the $error_{1\rightarrow N}$ is used as a final prediction accuracy for the evaluation. Fig.3 shows the final errors for AIC-lasso, BIC-lasso, Ranker-AIC-lasso, and Ranker-BIC-lasso.

It is shown from Fig.3 that all ensemble methodologies perform almost the same except with Banque Saudi Fransi predictors where AIC/BIC-Lasso performs slightly better than Ranker-AIC/BIC-Lasso . Therefore, the overall prediction for AIC/BIC-Lasso ensemble method outperforms Ranker-AIC/BIC-Lasso prediction accuracy but with slightly improvement. Since AIC-Lasso and BIC-Lasso ensemble methods preform better than Ranker-AIC/BIC-Lasso ensemble methods, it is important to choose the final ensemble method between AIC-Lasso and BIC-Lasso method as they perform the same in terms of their prediction accuracy. Since BIC tends to penalise complex models more heavily than AIC , and gives preference to simpler models in the search space [5], it is preferred over AIC-Lasso. This is because AIC-Lasso and BIC-Lasso give the same prediction accuracy and therefore, the simpler resultant model is favored. Hence, BIC-Lasso ensemble feature selection method is chosen as a final ensemble feature selection method to learn the causal-effect relationships between banks in the Saudi equities market. Fig.4 shows the learned causal-effect model using BIC-Lasso ensemble method.

To ensure good interpretation and visualisation we used different methods as follows:

The thickness of a an arrow means the amount of the directed causal relationship each predictor has in an effective bank. The sign illustrates the direction of the causality relationship. Thus, if it is '+' the interpretation is:

- When the price of a predictor increases, the effective bank price is increased.
- When the price of a predictor decreases, the effective bank price is decreased.

If the sign is '-', the interpretation is:

- When the price of a predictor increases, the effective bank price is decreased.
- When the price of a predictor bank decreases, the effective bank price is increased.

VI. Conclusion

This work shows machine learning of graphical models application in Saudi equities market. We have shown different developed ensemble feature selection methods to choose the best subset of predictors(causals) for a particular bank. We have given a comparison between the different proposed ensemble feature selection methods in terms of their prediction accuracy. The resultant graphical model from the best ensemble method, namely BIC-Lasso, shows the amount of causality each predictor has on a particular bank. The graphical model in Fig.4, shows that some predictors banks have different amount of causality on other banks and therefore, in the future work it is important to having inference on such learned model, to confirm the reality of causal-effect relationships inferred by our ensemble feature selection method.

Acknowledgment

The author would like to thank Saudi Stock Exchange market(Tadawul) for providing access to the historical data for Saudi equities market.

References

[1] Adel Aloraini, James Cussens, and Richard Birnie. Extending prostate cancer kegg pathways using machine learning of graphical models. In *Systemics and Informatics World Network,SIWN*, volume 10, pages 56–67, 2010.

[2] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 0(0):1–10, 2007.

[3] Dan Geiger and David Heckerman. Learning Gaussian networks. In *UAI*, volume 10, pages 235–243, 1994.

[4] Isabelle Guyon. *Practical Feature Selection: from Correlation to Causality*. IOS Press, January 16 2008.

[5] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer New York Inc., 2009.
[6] Tim C Hesterberg, Nam H Choi, Lukas Meier, and Chris Fraley. Least angle and l1 penalized regression: A review. *Statistics Surveys*, 2:61–93, 2008.
[7] Y. Hiemstra. Modeling structured nonlinear knowledge to predict stock market returns. *.R. Trippi (Ed.), Chaos and Nonlinear Dynamics in the Financial Markets: Theory, Evidence and Applications*, pages 163–175, 1995.
[8] K. Kim and I. Han. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Syst. Appl*, 19:125–132, 2000.
[9] K. Kim and I. Han. Financial time series forecasting using support vector machines. *Neurocomputing*, 55:307–319, 2003.
[10] KAZUHIRO KOHARA, TSUTOMU ISHIKAWA, YOSHIMI FUKUHARA, and YUKIHIRO NAKAMURA. Stock price prediction using prior knowledge and neural networks. *Intelligent Systems in Accounting, Finance and Management*, 6:12–22, 1997.
[11] Florian Markowetz and Rainer Spang. Inferring cellular networks - a review. *BMC Bioinformatics*, 8(S-6), 2007.
[12] Jie Peng, Pei Wang, Nengfeng Zhou, and Ji Zhu. Partial correlation estimation by joint sparse regression models. *American Statistical Association*, 104(486):735–746, 2009.
[13] Yvan Saeys, In aki Inza, and Pedro Larran aga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23:2507–2517, 2007.
[14] Robert J. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288, 1996.
[15] G. Zhang, B.E. Patuwo, and M.Y. Hu. Forecasting with artificial neural networksthe state of the art. *J. Forecasting*, 14:35–62, 1998.